Developing a Framework to Support Strategic Supply Chain Segmentation Decisions: a case study

Alexander A. Kharlamov

Bristol Business School, University of the West of England, Bristol, U.K.;

Luis Miguel D. F. Ferreira*

Department of Mechanical Engineering, Faculty of Sciences and Technology

University of Coimbra

Portugal

* Corresponding author e-mail: luis.ferreira@dem.uc.pt

Janet Godsell

WMG, IIPSI, The University of Warwick, Coventry, U.K.

Abstract

There is a huge opportunity in mining operational data in the supply chain (SC) to support strategic segmentation decisions. This research has the objective of developing a framework to support strategic supply chain segmentation decisions. This research is exploratory in nature, with the methodology based on action research combined with a single empirical study in a large Portuguese multinational company. A data-mining project, based on the CRISP-DM methodology, is adopted to develop the framework. The company had the strategic objective to move beyond a single make to order strategy towards a segmented SC strategy. By applying the framework, the most relevant criteria were identified (demand volume, demand variability, order corrections, delivery time window and delivery frequency). These were then used to identify four relevant segments each with a tailored SC strategy.

Keywords: supply chain management; segmentation; analytics; data mining.

Developing a Framework to Support Strategic Supply Chain Segmentation Decisions: a case study

There is a huge opportunity in mining operational data in the supply chain (SC) to support strategic segmentation decisions. This research has the objective of developing a framework to support strategic supply chain segmentation decisions. This research is exploratory in nature, with the methodology based on action research combined with a single empirical study in a large Portuguese multinational company. A data-mining project, based on the CRISP-DM methodology, is adopted to develop the product segmentation framework. The company had the strategic objective to move beyond a single make to order strategy towards a segmented SC strategy. By applying the framework, the most relevant criteria were identified (demand volume, demand variability, order corrections, delivery time window and delivery frequency). These were then used to identify four relevant segments each with a tailored SC strategy.

Keywords: supply chain management; segmentation; analytics; data mining

Introduction

1982 is widely recognised as the year that the term supply chain (SC) was first formally used by the Booz Hamilton consultants Oliver and Weber. The origins of the term 'Big Data' have proved more elusive despite concerted efforts1, and like the term Supply Chain, Big Data would appear to have practical origins. There is evidence to suggest (Diebold 2012) that it was first coined by Steven Mashey, the Chief Scientist of Silicon Graphics, in the early 1990s, and the first formal evidence of which is a presentation from 1998. Two of challenges previously identified by Steven Mashey, that have endured for almost 20 years, relate to having the required data, but then not being able to find and understand it and/or having it in the wrong place or form. In this way data analytics can help organisations to meet their objectives to better understand customers, improve products and services and improve internal efficiency (Kache and Seuring 2017; Arunachalam, Kumar, and Kawalek 2017; Bowers, Petrie, and Holcomb 2017; Papadopoulos et al. 2017; Roden et al. 2017; Ramanathan et al. 2017).

The field of analytics has been growing rapidly over the last 20 years. These types of approaches have been rapidly adopted in the fields of marketing, where outcome orientated approaches for data mining have become commonplace (Fan, Lau, and Zhao 2015; Xu, Frankwick, and Ramirez 2016). One of the most common examples of this is the way in which Dunnhumby has developed as a specialist analytics firm (Donnelly et al. 2015). It originates in gaining consumer insight from the huge amount of consumer data made available through the Tesco club card scheme. They took the basic process of consumer segmentation to a new level, as they were not constrained by analysing the data based on specific criteria, but were able to mine the data to identify the criteria that were most valuable to the objective they were trying to achieve. This inevitably was the identification of some form of consumer segmentation, to enable more targeted marketing efforts, to support sales growth.

---

1 See New York Times Column of Steve Lohr. https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/

The concept of SC segmentation shares a time line with Big Data. It started to gain traction in the late 1990s and was popularised by Fisher's (1997) Harvard Business Review article. The concept has been enduring but has typically sought to segment based on specific criteria. For simplicity of operationalisation many of these methods were reduced to two variables, e.g. volume and variability as in Godsell et al., (2011), however it is possible to go beyond this simplicity. Whilst previously multi-dimensional approaches (e.g. Christopher and Towill (2002), DWV3) proved too difficult to implement in practice, due to poor data availability and the internal analytical ability to process more than two variables, the situation is now very different. The problem of poor data availability and the analytical ability has been mostly overcome due to wide adoption of information systems in management. The problem of working with multidimensional data-sets is tamed with data-mining methods, which need to be incorporated in "pragmatic and practical solution" in order to be actionable. Therefore, the objective of this paper is to propose a pragmatic yet powerful approach to segmenting the SC using data-mining methods.

This paper presents an approach for strategic SC segmentation of Stock Keeping Units (SKU), that enables operational data to be considered, the most relevant criteria for segmentation to be identified, and segments operationalised. The paper firstly presents the rationale underpinning the proposed process for strategic SC segmentation. It then illustrates the development of the framework and its application on a case study from the Fast-Moving Consumer Goods (FMCG) industry. The paper draws to a close with a discussion of five key managerial insights that emerged from the case study.

Literature review

Segmenting Supply Chains

SC Segmentation started to gain prominence the mid-1990's. Under the auspices of the concept of strategic alignment (Gattorna and Walters 1996) the idea of matching supply chain strategy to different market segments was first introduced. Later, Naylor et al. (1999) laid the foundations for a debate about the ways in which lean and agile strategies can be combined.

The immaturity of SCM is widely recognised (Burgess, Singh, and Koroglu 2006; Ellram and Cooper 2014) and this is also reflected in adoption of segmentation in SCM (Protopappa-Sieke, Thonemann, and Thonemann 2017) Previous contributions (Table 1) relied heavily on using few pre-selected segmentation criteria mostly based on previous literature and focused on pre-determined number of segments. There has been limited consideration of how to fully operationalize the segments and the operationalization of the segments tended to be normative suggesting a prescribed solution rather than a more generalizable process or strategic framework (Lapide 2005b) that would lead to the development of a context specific solution.

At the current state, SC segmentation is normative and close-ended suggesting very specific criteria to be used. Different authors suggest different criteria depending on the context and purpose, and there are many possibilities suggested in the literature as listed in Table 1 which has been created using literature snowballing (Sayers 2007) and is in accordance with criteria mentioned throughout Protopappa-Sieke and Thonemann (2017). Literature Snowballing is a technique borrowed from systematic reviews as per Sayers (2007) and it is essentially working with the papers cited in the most recent literature identified and retrieving them, then lifting the citations from those until no more criteria have been identified.

Table 1 Criteria for Segmentation in SCM literature

Criteria / **Contributions**	Oliver and Webber (1982)	Fuller et al., (1993)	Gattorna and Walters (1996)	Fisher (1997)	Naylor *et al.* (1999)	Mason-Jones *et al.* (2000)	Childerhouse and Towill ((2000)	Lamming (2000)	Li & O'Brien (2001)	Christopher & Towill (2002)	Lee (2002)	Childerhouse *et al.* (2002)	Vitasek *et al.* (2003)	Aitken *et al.* (2003)	Bruce *et al.* (2004)	Payne and Peters (2004)	Holweg (2005)	Lovell *et al.* (2005)	Christopher *et al.* (2006)	Christopher *et al.* (2009)	Godsell *et al.* (2011)	Godsell *et al.* (2013)
Demand Variability					✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓
Lead time/Time window	✓		✓	✓		✓	✓	✓		✓	✓	✓		✓			✓		✓	✓		
Product life cycle				✓		✓	✓	✓		✓	✓	✓		✓	✓		✓	✓		✓		
Volume						✓		✓		✓	✓	✓	✓	✓		✓	✓			✓	✓	✓
Product Variety	✓			✓	✓		✓	✓		✓	✓	✓		✓			✓			✓		
Profit margin				✓		✓	✓		✓	✓	✓											
Demand Predictability				✓			✓		✓		✓								✓			✓
Flexibility	✓		✓					✓														
Point of product configuration								✓	✓								✓					
Responsiveness	✓					✓		✓									✓					
Customer expectations						✓		✓	✓								✓					
Reliability of supply	✓		✓					✓			✓											
Product complexity						✓		✓														
De-coupling point						✓											✓					
Frequency of delivery/orders			✓													✓						
Demand Pareto analysis													✓				✓					
Change over	✓										✓											
Range	✓							✓														
Stock out				✓			✓				✓											
Inventory costs				✓			✓				✓											
Minimum run size	✓																					
Quality problems											✓											
Obsolescence											✓											
Number of supply sources											✓											
Number of customers																✓						
Order line value																✓						
Substitutability of SKU																✓						
Price/Revenue		✓																				
Product Value Density																		✓				

Whilst SC segmentation is dominated by a priori approaches there is a slow shift to a more open-ended (i.e. data-driven) base selection and segment formation (Christopher et al. 2009; Godsell et al. 2013; Kharlamov, Ferreira, and Godsell 2013). Possibly one of the main reasons for this lack of sophistication in SC segmentation methods up until recently is the historical paucity of data in and between businesses (Christopher and Holweg 2011; Waller and Fawcett 2013). As operational data is becoming more and more consistent, accessible and complete (Davenport 2006; Marchand and Peppard 2013), it is no longer infeasible to use data-driven approaches in O&SCM. The progress towards post hoc and hybrid methods is referred as a 'more formal footing' by 'exploiting the "analytic corporation" IT capability' (Christopher et al. 2009, 260), combining data and intuition in deductive-abductive cycles to achieve meaningful SC segments (Godsell et al. 2013).

However, there is a general lack of practical detail on how to identify and formally select criteria for SC segmentation. Criteria for segmentation span a wide range of perspectives, such as supply characteristics (Naylor, Naim, and Berry 1999), product characteristics (Lovell, Saw, and Stimson 2005), demand volatility (Christopher, Peck, and Towill 2006), buying and returning behaviour (Pawar et al. 2016) or demand profile (Godsell et al. 2011). The identification as well as selection of criteria rely on wisdom and intuition and are only focused on exogenous dimensions. Once selected, segmentation criteria are always weighted equally. Similarly, segment identification is mostly done by boundary conditions and combinations of segmentation criteria, for example volume (low/high) against CV (low/high) that combined must form four segments (Godsell et al. 2011). To identify segments (cluster) most of the authors suggest a priori methods, in this case intuition based segments using lanes or levels (e.g. Vitasek et al. 2003; Payne and Peters 2004; Holweg 2005; Godsell et al. 2011), where "if too many levels are selected, then subsequent analysis becomes more complicated" (Christopher et al. 2009, 465).

To operationalize the SC segments, where tailored practices collected from literature are matched using both knowledge from O&SCM literature and managerial wisdom to judge the profiles of the previously obtained segments. Companies use a combination of practices for implementing appropriate supply chain strategies to gain competitive advantage (Dreyer et al. 2016; Datta 2017). Tailored practices (Lapide 2005a, 2005b; Godsell et al. 2011) are an alternative to best practices which are not necessary 'best' when combined with each other or with the context specifics. By assigning a set of tailored practices to each segment, a segment strategy is defined (e.g. Fuller et al. 1993; Lapide 2005a; Lapide 2005b; Godsell et al. 2011)

Mining data for strategic segmentation

The field of marketing has a long history of application of data mining methods for customer segmentation to gain insights (Wind and Cardozo 1974; Green 1977; Plank 1985; Henry and William 1986; Yankelovich and Meer 2006). These insights are then used to support the development of marketing strategies. Data mining methods solve a specific problem of understanding heterogeneity within large populations of customers and potential customers based on their individual characteristics (Franks 2009; Singleton 2009). SCM faces a similar problem considering its multiple suppliers, products and customers (e.g. Christopher et al., 2009; Christopher and Holweg 2011; Hofmann et al., 2013).

Data mining methods have had multiple applications in the context of SCM, for example, to reduce costs in manufacturing (Zhao, Johnsson, and He 2017), to optimise product configurations (Song and Kusiak 2009), to understand forecast behaviour (Altintas and Trick 2014) or for supplier selection in mass customisation (Ni, Xu, and Deng 2007). Leveraging the commonly abundant logistics data, data mining methods have also been applied to some extent in quality management in sustainable food SCs (Ting et al. 2014). However, application of data mining methods to support SC segmentation are not so common. This provides an opportunity to borrow methods from marketing to understand heterogeneity in the SCM context.

Segmentation process

Market segmentation in the marketing literature is mostly defined by its different segmentation methods (Wedel and Kamakura 2000). Initially segmentation processes started with pre-selected criteria and a pre-determined number of groups (a priori), evolving later into more sophisticated data-driven methods (post hoc or posteriori) when criteria and groups are determined on the basis of the data analysis (Wind 1978; Plank 1985; Wedel and Kamakura 2000; Kazbare, van Trijp, and Eskildsen

2010). Hybrid forms mix both a priori and post hoc methods. Methods for segmentation can also be classified, as suggested by Weld and Kamakura (2000) according to whether descriptive or predictive statistical methods are used. Descriptive methods deal with the associations across a single set of criteria while predictive methods analyse associations between two sets of criteria, independent and dependent.

The a priori segmentation approach starts with pre-conceived criteria for segmentation based on a manager's intuition and wisdom (Green 1977). It is known to be very effective for simple segmentation with few criteria, when the main goal is to obtain insights quickly about groups and about associations between criteria. However, it can easily get messy when using many criteria or many groups and is not very effective per se. In contrast, a post hoc segmentation approach, also known as a data driven approach, allows homogeneous groups to emerge from data based on similarities (clustering) supported by multivariate statistical methods. Post hoc methods are known to generate more powerful marketing programmes by using data-driven analysis to validate beliefs as well as to discover new insights and hidden patterns in the data (Clarke 2009; Martínez-López & Casillas 2009; Kazbare et al. 2010).

Clustering methods

Clustering applied to marketing was first reviewed by Punj and Stewart (1983) and is now a popular tool for post hoc descriptive segmentation. Clustering can be classified in two types: overlapping and non-overlapping. Non-overlapping is the most common and it can be hierarchical and non-hierarchical clustering. Hierarchical clustering starts with the highest number of clusters (one cluster per observation) and links them by successive steps based on a measure of proximity (e.g. Ward's method). A non-hierarchical method starts with a random initial division and through reorganisation of clusters optimises a certain criterion (e.g. K-means). Some argue that non-hierarchical methods are superior to hierarchical methods, as they are more robust to outliers and the presence of irrelevant attributes (Tsiptsis and Chorianopoulos 2009). A common problem for any clustering method is the determination of the number of clusters (e.g. Wedel and Kamakura 2000; Han and Kamber 2006).

Clustering has been applied in SC management before. For example, Ernst and Cohen (1990) describe a clustering procedure for production and inventory systems, based on similarity of operations, which offers flexibility regarding the criteria used as well as the number of final clusters. It is one of the few post-hoc approaches for differentiation in SC management that focuses on managing spare parts and also with an open-ended procedure (flexible in terms of criteria and final clusters), Duchessi et al., (1993) suggests a top-down approach analysing maintenance related data which at the time was not widely available. Canetta et al., (2005) focus on evaluating the most appropriate clustering methods for industrial databases, proposing a two-stage method. Moving towards prescriptive statistical analysis, demand categorization through clustering of short life-cycle technology products for capacity planning and capacity negotiation enabled insight into products with a limited operational history (Wu et al. 2006). Finally, one of the latest developments, relying on clustering in SC management, is the visual hierarchical clustering using self-organising maps to cluster SC entities (Chattopadhyay, Sengupta, and Sahay 2015). However, applications of clustering for SC management to support strategic SC segmentation decisions are scarce.

Methodology

In our paper, the development of the framework goes hand-in-hand with gaining important insights into strategic supply chain segmentation. Thus, researchers have to be directly involved in the process itself. Action research, which can be considered a special type of case study involving participation in the change, has therefore been selected as an appropriate research method (see, e.g. Westbrook 1995; Coughlan and Coghlan 2002; Hendry, Huang, and Stevenson 2013). Action research is appropriate when the focus of the research is on learning about changes and making improvements over time (Coughlan and Coghlan 2002). One of the main characteristics of action research is that researchers and clients actively collaborate through whole project, from the initial diagnosing to the final deployment (Bryman and Bell 2015). Action research generally looks at specific situation in real-time and uses analysis methods to understand the problem and suggest changes (Lee 1999). It is based on a set of stages, as specified by authors such as Meredith et al. (1989) or Coughlan and Coghlan (2009) and has been used in a series of recent studies (e.g., Ferreira et al. 2015; Silva et al. 2016).

Our research is based on a single case and on the application of the Cross Industry Standard Process for Data Mining, abbreviated as CRISP-DM, which is one of the most commonly used data-mining methodologies (Shearer et al. 2000). There are several potential alternatives for data mining processes proposed in the literature, namely KDD, CRISP-DM and SEMMA (Shearer et al. 2000; Mariscal, Marbán, and Fernández 2010; Zezzatti and Ochoa 2012; Shafique and Qaiser 2014). Comparatively CRISP-DM has been suggested as being more company-oriented than the very popular KDD and it is also more complete than SEMMA (e.g., Shafique and Qaiser 2014), which makes it the preferred process for this case study. CRISP-DM consists of six major phases and is cyclic in nature. The process starts with business understanding, followed by data-understanding, data preparation, modelling, evaluation and deployment. The link between the use of CRISP-DM for development of the framework and action research is illustrated in Figure 1.

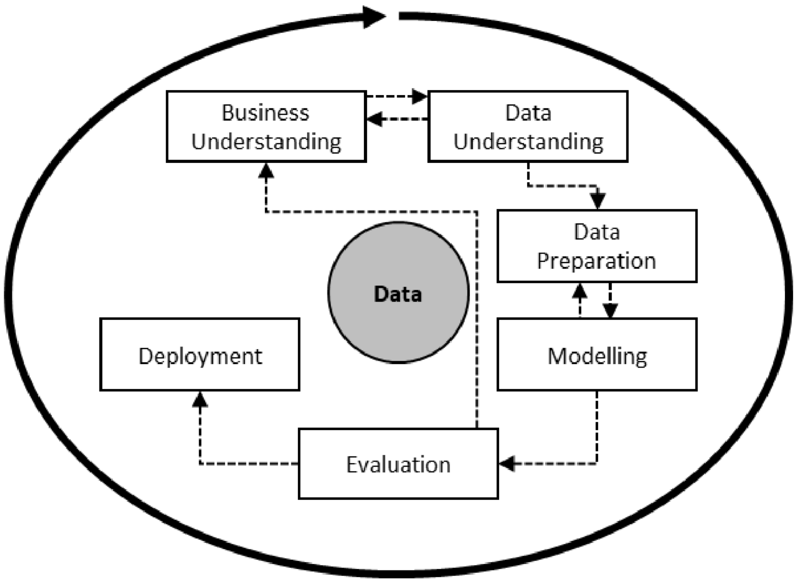

Figure 1 Development process of the framework based in the principles of action research CRISP-DM Cycle

Following Hendry, Huang, and Stevenson (2013), there are two main pitfalls associated with action research, which need to be avoided to ensure rigorous, high-quality outcomes. These pitfalls concern: (i) effective roles and relationships; and, (ii) appropriate data collection methods. Regarding (i), in this action research project, the first step was the formalisation of the project where the project objectives, duration, team composition and responsibilities were clearly defined. Moreover, the project sponsor was the global supply chain manager of the company, who was fully committed to the project and ensured the commitment of the remaining team members. Regarding (ii), a method to collect relevant data for the project was implemented. This method relies on the automatic reading of information extracted from the SAP information system, which allowed quantitative data for statistical analysis to be collected and reduced the probability of human data input errors.

Whilst a single application does not allow generalisation, it does contribute to the existing body of knowledge on SC segmentation both for academics and practitioners with potential of being used in other cases (e.g., Hofmann et al. 2013; Hersh 2014). Literature offers a wide range of papers based on Single studies and action research (Cagliano et al. 2005; Ferreira, Arantes, and Kharlamov 2015; Farooq and O'Brien 2015; Silva et al. 2016; Sunder 2016; Perona et al. 2016; Visintin et al. 2017; Visani and Bartolini 2018) and the use of case studies is welcome and encouraged (Childe 2011, 2017).

Action research study

Step 1: Business understanding

The data mining project begins with an understanding of the current situation of the firm and its business objective. The firm, referred to as FMCGCo, is a B2B company in the food industry, one of the major players in Europe with five production units. It specialises in processing fresh fruit purees that add flavour and/or colour to food products (e.g. dairy, beverages or pastries). On the supply side FMCGCo, is a large customer, with about 250 suppliers, the majority of which are fruit producers and cooperatives. Serving about one hundred clients, FMCGCo is a strategic supplier with very high level of integration and monitoring (e.g. special tracked food containers that plug directly into the customer's manufacturing line). Each product (SKU) is unique and is not shared with more than one client. Table 2 summarises the characteristics of the case.

Table 2: FMCGCo's characteristics

Characteristic	FMCGCo
Unit of analysis	Company (5 production units)
Time frame under analysis	2 years of operation
Industry	Food processing industry
Business relationship	B2B
Current manufacturing strategy	MTO
Experience	20+ years
Key product	Fruit-based additives and flavourings
Relationship with customers	Strategic supplier
Relationship with suppliers	Large customer (holds leverage)
Product type	Bespoke
Number of SKU	~ 1000
Number of suppliers	~ 250
Number of customers	~ 100
Average capacity utilisation	~75%
Capacity utilisation during high season	~95%
Seasonality	Peak between spring and summer
Demand volume trend	Positive growth
Geographic location	Europe & North of Africa

The project begins with unstructured interviews with managers and brainstorming. The aim is to understand what should be included in the exploratory analysis, understanding expectations and what the limitations, perceived problems and challenges are. Outputs are gathered through contact notes. The team included two researchers and representatives from the company. The company side included the global supply chain manager, production manager and internal control manager all dealing closely with the SAP and involved in production control, production planning and quality management. A total of 15 meetings were held focusing on data collection, development and validation. The average meeting took around two hours; all team members had equal weight in the decision process and the decisions were reached by consensus. When consensus on a decision was impossible to reach the decision was made based on the majority. All of the company sites had an integrated ERP system which enabled easy access to past data for analysis.

The company's overall operational goal and general directive is to be as responsive as possible. This attitude resonates in the global SC manager's own words as: "*…what the client wants, we deliver it. If our client wants star shaped Earliglow strawberries collected at 6 o'clock in the morning on the southern side of a mountain in a specific place of the world, we will engineer a process and find what is needed to deliver him that star shaped Earliglow strawberries…"*

The company works on a pure MTO basis (ETO for new products) which results in a high product variety (about one thousand SKUs). The product quality is very strict and being a perishable good the time window for delivery are often narrow. Being in the middle of the global supply chain, demand unpredictability is a great concern. Minor demand spikes downstream cause classical problems like bullwhip effect (Forrester, 1961), resulting in huge inventory build-ups upstream at the supplier's levels as well as "boom and bust" effect along SC (Sterman 2000). As managers suggested, this demand variability is the consequence of order and production batching, client's poor inventory management, lack of visibility and information sharing, as well as seasonality and final product marketing strategies (summer promotions). Most of the firm's operational efforts focus on coping with the volatile and unstable demand.

One of the problems and long-term concerns expressed by the managers is of production capacity. When discussing the manufacturing utilisation, demand peaks hit 96% utilisation during early summer and considering the current business growth the company was expected to start running out of capacity soon. In order to keep up the productivity without increasing manufacturing capacity, management stressed the importance of levelling the demand.

When discussing potential reasons for demand instability, managers complained about some clients repeatedly correcting their orders causing operational havoc. The team agreed that this could be measured using the ratio of corrected orders of a given SKU divided by the number of orders of the same SKU. Considering the nature of these order corrections, the main corrections were the delay of the delivery date (about 32%), followed by the anticipation of the delivery date (about 19%) and finally corrections to the ordered volume (13%). The team agreed that these corrections were completely driven by clients behaviour and that it causes significant disruptions to planning and manufacturing scheduling, often costing in lost raw materials and poor capacity utilisation.

The project team determined that the overall business goal is to identify the different SKU segments and their characteristics so that FMCGCo can deploy appropriate SC strategies. The SC management objective is to move beyond a one-size fits all strategy towards a segmented strategy.

The data-mining objective is therefore to cluster SKUs based on the operational data at the SKU level and to provide management with insights about the characteristics of each cluster informing the SC strategy.

Step 2: Data understanding and Data preparation

To facilitate the process a list of criteria previously used for differentiation in the SCM literature is presented (Table 3) and each criterion discussed.

Table 3: Criteria proposed for collection

Criteria		
Unavailable	**Available**	
Stock outs	Product life cycle	Change over*
Inventory costs	Minimum run size*	Substitutability of SKU
Quality problems	Point of product configuration	Range
Obsolescence	Responsiveness	Frequency of delivery**
Number of supply sources	Customer expectations	Demand Variability**
Order line value	Reliability of supply	Time window for delivery**
Flexibility	Product complexity*	Demand Volume**
Profit margin	Price/Revenue*	Order corrections ratio**
	Demand Predictability*	Frequency of orders
Not applicable to the case	Demand Pareto analysis	Product Variety*
Number of customers		
De-coupling point		
Product value density		

* Criteria considered for PCA analysis **Remaining criteria after PCA iterations

Some of the variables proposed are not available or not applicable. Stock outs, inventory costs, quality problems and others (see Table 3 for full list) were not available to the case. Other criteria such as profit margin was not obtained due to confidentiality. Reliability of supply was also removed due to very incomplete records. The number of customers and de-coupling points were not applicable to the case. Finally, one variable results from the brainstorming sessions with managers and was added to the list. This was into the frequency of order corrections that has great implications for production scheduling and overall efficiency. Order corrections is one of the drivers of "planning nervousness" (Andersen et al. 2014). Other variables such as product complexity derive from data in the list of materials for each SKU.

The demand history logged over two years at order level is aggregated weekly because both the expenditure and production scheduling are weekly as well. Other data was collected, including raw material consumption, raw material suppliers and inventory records, requirements for shipping of finished products, production stages for each SKU and other client data.

During data cleaning, some variables were completely homogeneous (e.g. point of product configuration), and others were very incomplete, and consequently dropped from further analysis as not having the ability to drive differentiation between observations. Finally, the set of criteria that made it to the dataset that was used in the PCA analysis included the following 11 criteria: Minimum run size, Change over, Order Corrections, Time window for delivery, Product complexity, Revenue, Demand Predictability, Demand Volume, Demand Variability, Product Variety and Frequency of deliveries.

The final dataset focused on products only, specifically stock keeping units (SKU). The motivation to segment at the SKU level is two-fold: first because SKUs were more varied than suppliers or clients, and, second, because classification of products has not received sufficient attention in the literature (e.g. Syntetos et al., 2009; Kampen et al., 2012).

Step 3: Modelling and Evaluation

Factor analysis is used to reduce the number of criteria for segmentation and to improve the clustering results. The measure of sampling adequacy suggested that the dataset is factorable (KMO = 0.630).

Principal component analysis (PCA) is used to select and simplify (reduce) into fewer components the criteria characterising the observations (Berthold and Hand 2007). PCA enables validation as well as discovery of previously unknown relationships between criteria and illuminates the structure of the dataset. PCA makes it possible to identify the data structure and reduce the number of criteria based on the correlations, i.e. relationships between criteria. This transformation automatically selects and reduces the variables making the extensive list of potential criteria for differentiation in the dataset more manageable. We originally start the PCA analysis including the 11 criteria as explained in the previous section. Non-relevant variables drop out of the analysis because they do not drive any differences. The 5 criteria (PCA input) are volume, demand variability, order corrections, delivery time window and delivery frequency. Those 5 criteria make up the 3 compound components (PCA output). Variety and revenue are introduced manually and used to describe the clusters rather than to form the clusters. The correlations between the selected criteria are listed in Table 4.

Table 4: Criteria correlation matrix (a=0.01, Sig. 1-tailed)

	Time Window for Delivery	**Order Correction**	**Demand Volume**	**Demand Variability**
Order Corrections	-0.030 (0.184)			
Demand Volume	-0.086 (0.004)	-0.022 (0.255)		
Demand Variability	0.239 (0.000)	0.028 (0.200)	-0.150 (0.000)	
Frequency of deliveries	-0.107 (0.001)	-0.185 (0.000)	0.527 (0.000)	-0.159 (0.000)

As shown in Table 4, in this specific case Demand Variability is significantly correlated to the Time Window for Delivery as well as to Demand Volume. Frequency of Deliveries is significantly correlated to the Time Window for Delivery, Order Corrections, Demand Volume and Demand Variability. The strongest correlation is between Volume and Frequency of Delivery (weeks with delivery) while the second strongest correlation is between Demand Variability and Time Window for Delivery.

Table 5: Total Variance Explained (before rotation)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	Variance %	Cumul. %	Total	Variance %	Cumul. %
1	1.700	33.994	33.994	1.700	33.994	33.994
2	1.135	22.699	56.693	1.135	22.699	56.693
3	0.965	19.300	75.993	0.965	19.300	**75.993**
4	0.753	15.063	91.057			
5	0.447	8.943	100.000			

Determination of the number of components to extract is based on the total variance accounted for (Table 5). Rules for component extraction are ad-hoc, with no theoretical justification and rely mostly on judgement and practicality. The actual 75.993% is considered acceptable, on the basis of three principal components (Berthold and Hand 2007). Orthogonal rotation eases the interpretation of the components, and produces uncorrelated components. The rotation method applied in this study was the **“**Varimax” with Kaiser Normalization, which converged in four iterations. After orthogonal rotation, the total variance is redistributed and each value accounts for the variance of the original variance contained in each component.

Table 6: Rotated component matrix (loadings less than 30% omitted)

Criteria	Component			Communalities	
	1	2	3	Initial	Extraction
Demand Volume	0.888			1.000	0.656
Frequency of deliveries	0.846			1.000	0.972
Time Window for Delivery		0.806		1.000	0.802
Demand Variability		0.761		1.000	0.602
Order Corrections			0.984	1.000	0.767

The outputs from the PCA analysis is three components that are a simplification of the original criteria (Table 6) to be used further for Clustering. Component 1 is composed of Demand Volume and Frequency of Deliveries and represents variances of 89% for volume and 85% of weeks with delivery (frequency of delivery) both positively correlated, with 0.527 significant correlation meaning that greater Demand Volume is associated with greater Frequency of Deliveries. Component 2 (which incorporates Time Window for Delivery & Demand Variability) accounts for 90% of Time Window for Delivery variance as well as Demand Variability variance, both positively correlated (0.237), meaning that a longer time window for delivery tends go together with greater Demand Variability. Finally, Component 3 includes only Order Corrections and reflects the client's inventory management and planning performance. The list of commonalities (Table 6) for each original dimension after the extraction ranges from 0.602 (the worst) to 0.972 (the best). The solution is accepted as satisfactory because extractions above 0.500 are considered acceptable (Berthold and Hand 2007).

Table 7: Product clusters profiled against the selected criteria

	Criteria	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster profile	Variety	21%	54%	18%	7%
	Revenue	63%	26%	8%	3%
	Demand Volume	74%	10%	8%	8%
	Demand Variability	Stable	n.s.	n.s.	Unstable
	Order Corrections	Average	V. low	High	Low
	Delivery Time Window	V. short	Short	Short	V. long
	Delivery Frequency	Frequent	Sporadic	V. Rare	Average

The chosen cluster analysis method is Ward's hierarchical method based on the squared Euclidean distance between products using the 3 orthogonal components coming out of the PCA analysis. The choice of clustering method is based on dataset size and the time needed to obtain a solution. For relatively small datasets, hierarchical methods provide a level of detail that can easily be understood by managers as it gradually aggregates SKUs based on their similarity at different levels of aggregation. To evaluate the number of clusters to extract the dendrogram was used (see figure 2). The final clustering and respective profiles against the selected criteria are summarised in Table 7.

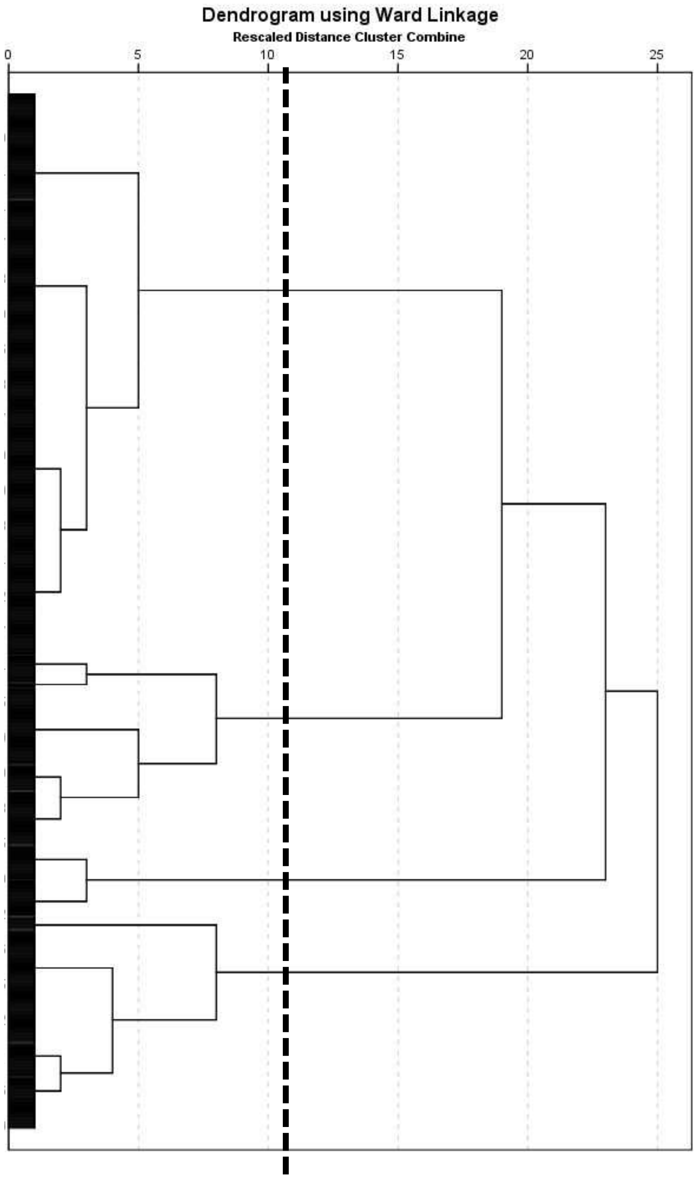

Figure 2 Clustering dendrogram

The first cluster is characterised by products the shortest Delivery Time Window and high Delivery Frequency, the latter associated with high Demand Volume and the lowest Demand Variability. The second cluster is characterised by products the fewest Order Corrections as well as low Demand Volume per product. The third cluster contains products delivered rarely (low Delivery Frequency) and low average monthly Demand Volume with the highest Order Correction. The final fourth cluster contains products with the longest average Delivery Time Window and the highest Demand Variability. Using the data from Table 7, a cumulative distribution is plotted in Figure 3 to represent the relationship between product variety and revenue for the clusters.

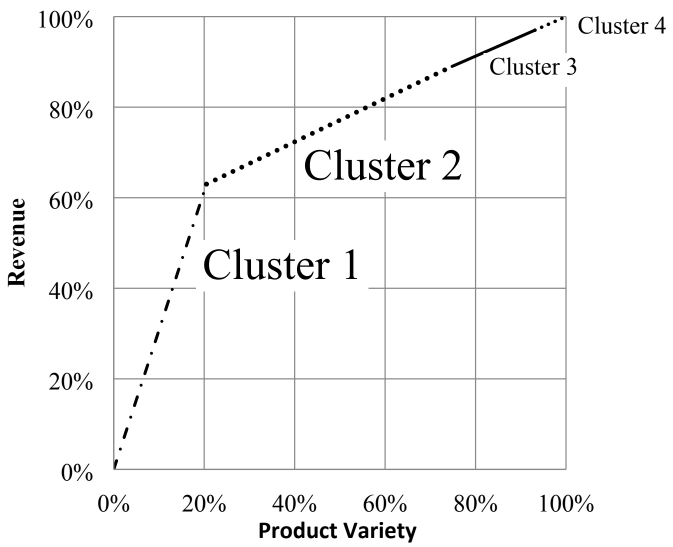

Figure 3: Revenue versus Variety - cumulative analysis of clusters

Step 4: Deployment

The deployment stage focuses on two directions, operational priority (Fine 1998) and planning approach (Stadtler 2005). Knowing the segments characteristics, we match SC practices with segments in order to gain competitive advantage (Dreyer et al. 2016; Datta 2017). The main insights span two main levels, from the criteria analysis (Table 6) and from cluster analysis with the respective profiles (Table 7), together with the cumulative distribution of product variety and total revenue per cluster illustrated in Figure 3.

Considering the criteria analysis, as expected the Demand Volume is significantly correlated with Frequency of Deliveries, showing that greater average demand is also delivered more frequently in similarly sized batches. Regarding the positive relationship between Time Window for Delivery and Demand Variability, greater flexibility in delivering the order was associated to greater instability in demand volumes, which was caused by the greater uncertainty associated with longer time-horizons when planning.

Focusing on the solution with four clusters, several insights can be drawn. The majority of the demand volume (74%) is stable/predictable (Cluster 1) and could be made to forecast (MTF) and forecast statistically (Syntetos et al. 2016). Considering cumulative variety and revenue in Figure 3 the main insight is that only 21% of the products account for 63% of the revenue meaning that most of the demand management can be automated and optimised. Cluster 2 suggests the need to streamline processes and reduce variety (Ramdas 2003), possibly through techniques such as postponement (Pagh and Cooper 1998).

Cluster 3 is a problematic segment due to order changes, so some customers could be approached with the suggestion that they implement vendor managed inventory (VMI) and collaborative planning, which would increase their visibility (Lapide 2008). High Order Corrections is most likely indicative of client's poor inventory management. Order Corrections introduce variability/unpredictability that is not reflected in the Demand Variability but contributes considerably to capacity utilisation peaks due to frequent rescheduling in the high season. The main reason for the relative stability observed in some of the clusters is weekly batching and great efforts to meet all order quantity corrections in an overall stable demand.

Finally, Cluster 4 brings together the few products that could potentially be discontinued and substituted by other more common products, because of their unstable and sporadic demand. These should be kept on an MTO basis (Tenhiälä and Ketokivi 2012). The summary of recommended SCM planning practices for each cluster is listed in Table 8, drawing on unstructured interviews and discussion of the outputs with the managers.

Table 8: Illustrative matching of SCM planning practices with product clusters

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Operational Priority	Reduce operational cost (Fisher and Raman 1996)	✓			
	Reduce Variety & Stream line processes (Ramdas 2003)		✓		
	Negotiate end of life / substitution (Aitken, Childerhouse, and Towill 2003)			✓	✓
Planning Approach	MTF - Forecast statistically (Syntetos et al. 2016)	✓			
	Collaborative planning/VMI (Lapide 2008)			✓	
	MTO (Tenhiälä and Ketokivi 2012)		✓		✓

Insights from the case study

The potential benefits of 'borrowing' from the field of marketing, to mine a blend of structured and unstructured data, against a strategic SC objective have been demonstrated through the FMCGCo case. FMCGCo has the strategic objective to move beyond a single MTO strategy towards a segmented SC strategy and by following the data-mining methodology, the five most relevant criteria (demand volume, demand variability, order corrections, delivery time window and delivery frequency) were identified. These were then used to identify the relevant segments, that require different SC responses. While the primary insight from this study is the empirically tested approach for data mining to support strategic SC decision making, five further specific takeaways for industry emerged.

Finding the Known Unknowns

One of the strengths of this approach to data mining is that it helps to identify the 'known unknowns'. Popularised by Donald Rumsfeld in a Department of Defence briefing (February 12th 2002), 'Known UnKnowns' are things that we know we do not know. Four of the five criteria that were identified are part of Christopher and Towill's (2002) DVW^3 approach. The fifth dimension, order correction, has no previous mention in the literature. It was identified in Step 1, through conversation with the management of FMCGCo. It was part of their business that was known to them, as they had to deal with the impact of changing orders all the time. The significance of this factor was ‘unknown’ to them and only identified through the data mining.

Turns insight into SC practices

A critical outcome of any form of clustering or segmentation is that it drives differentiated practices within the business. Identifying different segments enables the development of specific marketing programmes for the different segments. Planning is the glue that holds an SC together, and it is not surprising that the primary differentiator between the identified segments in terms of operationalisation is the planning process combined with the operational priority. Building on the first insight, it is ‘known’ that different SKUs may require different processes, but what can be difficult to determine (i.e. ‘unknown’) is which SKUs require which process. This method is very effective in identifying the clusters of SKUs that require different planning approaches.

The need to combine data with managerial insight

Managerial insight is critical throughout the process, and particularly in Steps 1 and 3. Although the manager may not have recognised its importance, without the conversation order correction would never have been identified as a potential dimension in Step 1. It is therefore important to ensure that in the first stage of the process, a detailed understanding of the business, its challenges and opportunities are understood to ensure that the initial list of dimensions is as broad as possible. At this stage, it may be useful to involve someone from outside the business, who can look at the business with a fresh pair of eyes so that critical aspects of the SC operation (e.g. customers continually changing their orders) is not overlooked. The operationalisation of the clusters in Step 3 would not be possible without the input of the SC team, who have the SC knowledge to understand the operational priorities and how they can be operationalised in practice.

Understand the relative importance of different segments

George Orwell (1945, p.40) in his novel Animal Farm famously said “All animals are equal, but some animals are more equal than others". The same can be said for SC clusters. As illustrated in Figure 3, Clusters 1 and 2 accounted for over 80% of the revenue, and almost 80% of the variety of the FMCGCo portfolio. These segments were the 'cash cows' for the business. To enable these important segments to be serviced as efficiently as possible, the SKUs in these clusters need to be recognised and given appropriate management attention.

Increasing importance of the SC analyst

While many of the major consultancies are starting to develop their SC capability more formally, and specialist SC analytic companies are still in an embryonic field. As this study has identified, the potential benefits to a company are huge. The question is not about whether to deploy SC analytics in a business, but how best to deploy the approaches. Depending on the size, maturity, and the appetite to put data driven approaches at the heart of the business, companies may choose to develop this as an in-house capability or through engaging the services of a third party.

Implementation and implications for practice

After the deployment stage the firm spent the following months implementing the insights of this study. The follow up evaluation of benefits of this study was not scoped, but the team made a qualitative assessment of the implementation. It was common agreement that we cannot fully attribute all the improvements in firms operational performance observed over the following months to the changes made based on this study, however, the team believes that it had a crucial role.

Part of the produce is now MTF rather than MTO. A new programme started aiming at establishing VMI and collaborative planning with its main client. Some of the SKUs were discontinued and substitutes common to several clients were developed and agreed with clients motivated by a lower price due to the economy of scale. Following these changes, the firm observed a noticeable reduction in operational cost and increase in productivity though better capacity utilisation due to demand levelling and reduction of variability. Some of the "bad clients" were identified and contacted with suggestions to address the order corrections problem.

The global SC manager says that: "… *we are still offering the best in class customisation and most of what we produce is MTO because that is still the core of what we do. But we no longer do this for every single client or product, part is standardized which allows us to have lower costs and we share this benefit with our clients. It's a clear win-win for all!*" adding that "…*we realise that by trying to respond to any customer demand we were underperforming. Now we are looking for a healthy balance between being responsive and being efficien*t."

Regardless of all the positive dynamic observed in a relatively short timeframe, the team realises that maximizing the potential of SC segmentation will take a long time, effort and commitment. The close involvement of both researchers and practitioners though the action research cycle was also praised due to its agility and efficiency and benefitted all parties involved.

Conclusions

The proposed framework aims at integrating data mining methods into supply chain segmentation. By applying the CRISP-DM Cycle and Action Research we identify the five most relevant criteria (demand volume, demand variability, order corrections, delivery time window and delivery frequency). We use these criteria to identify the four segments that require different SC responses. The action research used for this study contributes to developing new knowledge as well as achieving practical results. The main advantages of action research is the close involvement of both researchers and practitioners in a common effort. The cyclical process of planning, acting and evaluating is the core of action research methodology, enabling a much faster research process compared to other, more traditional and detached methods. We identify the following five key takeaways: finding the Known Unknowns, turns insight into SC practices, the need to combine data with managerial insight, understand the relative importance of different clusters and increasing importance of the SC analyst.

The framework is innovative due to the application of data mining methods in the SCM context. The framework is based on The Cross Industry Standard Process for Data Mining. This makes our framework scalable and compatible with much larger portfolios. The main characteristics of the framework is its open-ended nature, expansibility and adaptability. The use of exploratory data mining methods not only allows validation but also discovering new insights from data. The proposed framework can be used as a regular diagnostic check tool to constantly re-assess segments based on the most recent demand history partly solving the problem of demand dynamics.

The framework provides analytical means to select criteria driving the differences between products or clients. Boundary conditions between categories while previously arbitrary and prescriptive, are now data-driven, i.e. naturally defined by the data structure assisted by managerial insight. The discovery of previously unknown aspects about the SCM context lead to better understanding of the supply, product portfolio and demand which may reduce the cost to serve. Finally, additionally to a retrospective analysis of operational data, the framework is compatible with real-time analysis if automatized and fed with real-time data, allowing the monitoring of changes.

The empirical application focused on products, but the framework has the potential to include both suppliers and client's data. Further development should focus on the application of the framework on other cases, for example in on other industries beyond FMCG context, and potentially the adoption of more volume-efficient clustering methods (e.g. K-means clustering).

References

Aitken, James, Paul Childerhouse, and Denis Towill. 2003. "The Impact of Product Life Cycle on Supply Chain Strategy." *International Journal of Production Economics* 85 (2): 127–40.

Altintas, Nihat, and Michael Trick. 2014. "A Data Mining Approach to Forecast Behavior." *Annals of Operations Research* 216 (1): 3–22.

Andersen, Ann-Louise, Nicolai Præstholm, Kjeld Nielsen, and Thomas Ditlev Brunø. 2014. "Planning Nervousness in Product Segmentation: Literature Review and Research Agenda." In *IFIP International Conference on Advances in Production Management Systems*, 403–10. Springer.

Arunachalam, Deepak, Niraj Kumar, and John Paul Kawalek. 2017. "Understanding Big Data Analytics Capabilities in Supply Chain Management: Unravelling the Issues, Challenges and Implications for Practice." *Transportation Research Part E: Logistics and Transportation Review*, May. Pergamon.

Berthold, Michael, and David J. Hand. 2007. *Intelligent Data Analysis*. 2nd ed. Springer.

Bonoma, Thomas V., and Benson P. Shapiro. 1984. "Evaluating Market Segmentation Approaches." *Industrial Marketing Management* 13 (4): 257–68.

Bowers, Melissa R, Adam Petrie, and Mary C Holcomb. 2017. "Unleashing the Potential of Supply Chain Analytics." *MIT Sloan Management Review* 59 (1). Massachusetts Institute of Technology, Cambridge, MA: 14.

Bruce, Margaret, Lucy Daly, and Neil Towers. 2004. "Lean or Agile: A Solution for Supply Chain Management in the Textiles and Clothing Industry?" *International Journal of Operations & Production Management* 24 (2): 151–70.

Bryman, Alan, and Emma Bell. 2015. *Business Research Methods*. Oxford University Press, USA.

Burgess, Kevin, Prakash J. Singh, and Rana Koroglu. 2006. "Supply Chain Management: A Structured Literature Review and Implications for Future Research." *International Journal of Operations & Production Management* 26 (7): 703–29.

Cagliano, R, F Caniato *, M Corso, and G Spina. 2005. "Collaborative Improvement in the Extended Manufacturing Enterprise: Lessons from an Action Research Process." *Production Planning & Control* 16 (4), 345–55.

Canetta, L., N. Cheikhrouhou, and R. Glardon. 2005. "Applying Two-Stage SOM-Based Clustering Approaches to Industrial Data Analysis." *Production Planning & Control* 16 (8), 774–84.

Chattopadhyay, Manojit, Sourav Sengupta, and B.S. Sahay. 2015. "Visual Hierarchical Clustering of Supply Chain Using Growing Hierarchical Self-Organising Map Algorithm." *International Journal of Production Research* 7543 (October 2015): 1–20.

Childe, Stephen. 2011. "Case Studies in Operations Management." *Production Planning & Control* 22 (2). 107.

———. 2017. "Case Studies in the Management of Operations." *Production Planning & Control* 28 (1). 1.

Childerhouse, Paul, James Aitken, and Denis R. Towill. 2002. "Analysis and Design of Focused Demand Chains." *Journal of Operations Management* 20 (6): 675–89.

Childerhouse, Paul, and Denis Towill. 2000. "Engineering Supply Chains to Match Customer Requirements." *Logistics Information Management* 13 (6): 337–46.

Christopher, Martin, and Matthias Holweg. 2011. "'Supply Chain 2.0': Managing Supply Chains in the Era of Turbulence." *International Journal of Physical Distribution & Logistics Management* 41 (1): 63–82.

Christopher, Martin, Helen Peck, and Denis Towill. 2006. "A Taxonomy for Selecting Global Supply Chain Strategies." *The International Journal of Logistics Management* 17 (2): 277–87.

Christopher, Martin, and Denis R. Towill. 2002. "Developing Market Specific Supply Chain Strategies." *International Journal of Logistics Management* 13 (1): 1–14.

Christopher, Martin, Denis R. Towill, James Aitken, and Paul Childerhouse. 2009. "Value Stream Classification." *Journal of Manufacturing Technology Management* 20 (4): 460–74.

Clarke, Ann H∅jbjerg. 2009. "Bridging Industrial Segmentation Theory and Practice." *Journal of Business-to-Business Marketing* 16 (4): 343–73.

Datta, Partha Priya. 2017. "Enhancing Competitive Advantage by Constructing Supply Chains to Achieve Superior Performance." *Production Planning & Control* 28 (1), 57–74.

Davenport, Thomas H. 2006. "Competing on Analytics." *Harvard Business Review*, no. January: 1–10.

Diebold, Francis. 2012. "On the Origin (s) and Development of the Term Big Data." *Penn Institute for Economic Research (PIER), Research Paper Series*.

Donnelly, Christina, Geoff Simmons, Gillian Armstrong, and Andrew Fearne. 2015. "Digital Loyalty Card 'Big Data'and Small Business Marketing: Formal versus Informal or Complementary?" *International Small Business Journal* 33 (4). SAGE Publications Sage UK: London, England: 422–42.

Dreyer, Heidi C, Jan O Strandhagen, Hans-Henrik Hvolby, Anita Romsdal, and Erlend Alfnes. 2016. "Supply Chain Strategies for Speciality Foods: A Norwegian Case Study." *Production Planning & Control* 27 (11), 878–93.

Duchessi, Peter, Giri Kumar Tayi, and Joshua B Levy. 1993. "A Conceptual Approach for Managing of Spare Parts." *International Journal of Physical Distribution & Materials Management* 18 (5): 8–15.

Ellram, Lisa M, and Martha C Cooper. 2014. "Supply Chain Management: It's All about the Journey, Not the Destination." *Journal of Supply Chain Management* 50 (1). Wiley Online Library: 8–20.

Ernst, Ricardo, and Morris a. Cohen. 1990. "Operations Related Groups (ORGs): A Clustering Procedure for Production/Inventory Systems." *Journal of Operations Management* 9 (4): 574–98.

Fan, Shaokun, Raymond Y.K. Lau, and J. Leon Zhao. 2015. "Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix." *Big Data Research* 2 (1). Elsevier: 28–32.

Farooq, Sami, and Chris O'Brien. 2015. "An Action Research Methodology for Manufacturing Technology Selection: A Supply Chain Perspective." *Production Planning & Control* 26 (6), 467–88.

Ferreira, Luís Miguel D F, Amílcar Arantes, and Alexander A Kharlamov. 2015. "Development of a Purchasing Portfolio Model for the Construction Industry: An Empirical Study." *Production*

Planning & Control, 26 (5), 377–92.

Fine, C H. 1998. “Winning Industry Control in the Age of Temporary Advantage. Sloan School of Management.” MIT Addison-Wesley, Boston.

Fisher, M., and a. Raman. 1996. “Reducing the Cost of Demand Uncertainty Through Accurate Response to Early Sales.” *Operations Research*.

Fisher, ML. 1997. “What Is the Right Supply Chain for Your Product?” *Harvard Business Review* 75 (2): 105–16.

Franks, Bill. 2009. “Data Warehousing and Consumer Analytics in the Era of Customer Engagement – An Interview with Bill Franks of Teradata.” *Journal of Digital Asset Management* 5 (6): 352–63.

Fuller, JB, J O’Conor, and R Rawlinson. 1993. “Tailored Logistics: The next Advantage.” *Harvard Business Review* 71 (3): 87–93.

Gattorna, J.L., and D.W. Walters. 1996. *Managing the Supply Chain – A Strategic Perspective*. London: Macmillan.

Godsell, Janet, Thomas Diefenbach, Chris Clemmow, Denis Towill, and Martin Christopher. 2011. “Enabling Supply Chain Segmentation through Demand Profiling.” *International Journal of Physical Distribution & Logistics Management* 41 (3): 296–314.

Godsell, Janet, Alexander Kharlamov, Sampada Vasishta, and John Burdett. 2013. “Challenging the Volume-Variability Paradigm : An Empirical Investigation in a Global Beverage Company.” In *Proceedings - 2013 20th International Annual EurOMA Conference*. Dublin.

Green, PE. 1977. “A New Approach to Market Segmentation.” *Business Horizons* 20 (1): 61–73.

Han, Jiawei, and Micheline Kamber. 2006. *Data Mining: Concepts and Techniques*. 2nd ed. Elsevier Inc.

Henry, J, and F William. 1986. “A Theory of Market Segmentation.”

Hersh, Mark. 2014. “Supply Chain Segmentation : Mastering Complexity.”

Hofmann, Erik, Patrick Beck, and Erik Füger. 2013. *The Supply Chain Differentiation Guide*. London: Springer.

Holweg, Matthias. 2005. “The Three Dimensions of Responsiveness.” *International Journal of Operations & Production Management* 25 (7): 603–22.

Kache, Florian, and Stefan Seuring. 2017. “Challenges and Opportunities of Digital Information at the Intersection of Big Data Analytics and Supply Chain Management.” *International Journal of Operations & Production Management* 37 (1). Emerald: 10–36.

Kampen, Tim J. Van, Renzo Akkerman, and Dirk Pieter Van Donk. 2012. “SKU Classification: A Literature Review and Conceptual Framework.” *International Journal of Operations & Production Management* 32 (7): 850–76.

Kazbare, Laura, Hans C.M. van Trijp, and Jacob Kjær Eskildsen. 2010. “A-Priori and Post-Hoc Segmentation in the Design of Healthy Eating Campaigns.” *Journal of Marketing Communications* 16 (1–2): 21–45.

Kharlamov, AlexanderA., LuisMiguelD.F. Miguel D. F. Ferreira, and Janet Godsell. 2013. “The Power of Analytical Approaches Towards the Development of Differentiated Supply Chain Strategies: Case Study.” In *Advances in Sustainable and Competitive Manufacturing Systems SE - 100*, edited by Américo Azevedo, 1223–35. Lecture Notes in Mechanical Engineering. Springer International Publishing.

Lamming, Richard, Thomas Johnsen, Jurong Zheng, and Christine Harland. 2000. “An Initial Classification of Supply Networks.” *International Journal of Operations & Production Management* 20 (6): 675–91.

Lapide, L. 2005a. “Benchmarking Best Practices.” *The Journal of Business Forecasting* 24 (4): 29–32.

———. 2005b. “The Four Habits of Highly Effective Supply Chains.” *Harvard Business Review* 1 (3): 1–6.

Lapide, Larry. 2008. “Use VMI to Improve Forecasting.” *Journal of Business Forecasting*, 28–31.

Lee, HL. 2002. “Aligning Supply Chain Strategies with Product Uncertainties.” *California Management Review* 44 (3): 105–19.

Lee, Thomas W. 1999. *Using Qualitative Methods in Organizational Research*. Sage.

Li, D., and C. O’Brien. 2001. “A Quantitative Analysis of Relationships between Product Types and Supply Chain Strategies.” *International Journal of Production Economics* 73 (1): 29–39.

Lovell, Antony, Richard Saw, and Jennifer Stimson. 2005. “Product Value-Density: Managing Diversity through Supply Chain Segmentation.” *The International Journal of Logistics Management* 16 (1): 142–58.

Marchand, Donald A., and Joe Peppard. 2013. “Why IT Fumbles Analytics.” *Harvard Business Review*, no. January-February: 104–12.

Mariscal, Gonzalo, Óscar Marbán, and Covadonga Fernández. 2010. “A Survey of Data Mining and Knowledge Discovery Process Models and Methodologies.” *The Knowledge Engineering Review* 25 (02): 137–66.

Martínez-López, Francisco J, and Jorge Casillas. 2009. “Marketing Intelligent Systems for Consumer

Behaviour Modelling by a Descriptive Induction Approach Based on Genetic Fuzzy Systems." *Industrial Marketing Management* 38 (7). New York: Elsevier Sequoia S.A.: 714.

Mason-Jones, Rachel, Ben Naylor, and D.R. Towill. 2000. "Lean, Agile or Leagile? Matching Your Supply Chain to the Marketplace." *International Journal of Production Research* 38 (17): 4061–70.

Naylor, J Ben, Mohamed M Naim, and Danny Berry. 1999. "Leagility: Integrating the Lean and Agile Manufacturing Paradigms in the Total Supply Chain." *International Journal of Production Economics* 62 (1–2): 107–18.

Ni, M, X Xu, and S Deng. 2007. "Extended QFD and Data-Mining-Based Methods for Supplier Selection in Mass Customization." *International Journal of Computer Integrated Manufacturing* 20 (2–3). Taylor & Francis: 280–91.

Oliver, R. K., and M. D. Weber. 1982. "Supply-Chain Management: Logistics Catches Up With Strategy." In *Logistics: The Strategic Issues*, edited by M. G. Christopher, 63–75. London: Chapman & Hall.

Pagh, JD, and MC Cooper. 1998. "Supply Chain Postponement and Speculation Strategies: How to Choose the Right Strategy." *Journal of Business Logistics* 19 (2): 13–33.

Papadopoulos, Thanos, Angappa Gunasekaran, Rameshwar Dubey, and Samuel Fosso Wamba. 2017. "Big Data and Analytics in Operations and Supply Chain Management: Managerial Aspects and Practical Challenges." *Production Planning & Control* 28 (11–12). Taylor & Francis: 873–76.

Pawar, Kulwant S, Helen Rogers, Andrew Potter, and Mohamed Naim. 2016. *Developments in Logistics and Supply Chain Management: Past, Present and Future*. Springer.

Payne, Tim, and Melvyn J. Peters. 2004. "What Is the Right Supply Chain For Your Products?" *The International Journal of Logistics Management* 15 (2): 77–92.

Perona, Marco, Nicola Saccani, Stefano Bonetti, and Andrea Bacchetti. 2016. "Manufacturing Lead Time Shortening and Stabilisation by Means of Workload Control: An Action Research and a New Method." *Production Planning & Control* 27 (7–8). Taylor & Francis: 660–70.

Plank, Richard E. 1985. "A Critical Review of Industrial Market Segmentation." *Industrial Marketing Management* 14 (2): 79–91.

Protopappa-Sieke, Margarita, Ulrich Thonemann, and Ulrich W Thonemann. 2017. *Supply Chain Segmentation*. Springer.

Punj, Girish, and David Stewart. 1983. "Cluster Analysis In Marketing Research: Review and Suggestions for Application." *Journal of Marketing Research* 20 (2): 134–48.

Ramanathan, Ramakrishnan, Elly Philpott, Yanqing Duan, and Guangming Cao. 2017. "Adoption of

Business Analytics and Impact on Performance: A Qualitative Study in Retail." *Production Planning & Control* 28 (11–12). Taylor & Francis: 985–98.

Ramdas, Kamalini. 2003. "Managing Product Variety: An Integrative Review and Research Directions." *Production and Operations Management* 12 (1): 79–101.

Roden, S, A Nucciarelli, F Li, and G Graham. 2017. "Big Data and the Transformation of Operations Models: A Framework and a New Research Agenda." *Production Planning & Control* 28 (11–12). Taylor & Francis: 929–44.

Sayers, Adrian. 2007. "Tips and Tricks in Performing a Systematic Review." *The British Journal of General Practice* 57 (542). Oxford University Press: 759.

Shafique, Umair, and Haseeb Qaiser. 2014. "A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA)." *International Journal of Innovation and Scientific Research ISSN* 12 (1): 2351–8014.

Shearer, Colin, Hugh J Watson, Daryl G Grecich, Larissa Moss, Sid Adelman, Katherine Hammer, and Stacey A Herdlein. 2000. "The CRISP-DM Model: The New Blueprint for Data Mining." *Journal of Data Warehousing* 5 (4): 13–22.

Silva, Cristovao, Luis Miguel Ferreira, Matthias Thürer, and Mark Stevenson. 2016. "Improving the Logistics of a Constant Order-Cycle Kanban System." *Production Planning & Control* 27 (7–8). 650–59.

Singleton, Alex D. 2009. "Data Mining Course Choice Sets and Behaviours for Target Marketing of Higher Education." *Journal of Targeting, Measurement and Analysis for Marketing* 17 (3). London: Palgrave Macmillan: 157–70.

Song, Z., and A. Kusiak. 2009. "Optimising Product Configurations with a Data-Mining Approach." *International Journal of Production Research* 47 (7): 1733–51.

Stadtler, Hartmut. 2005. "Supply Chain Management and Advanced Planning—Basics, Overview and Challenges." *European Journal of Operational Research* 163 (3): 575–88.

Sterman, J D. 2000. *Systems Thinking and Modeling for a Complex World. Management*, Irwin/McGraw-Hill.

Sunder M, Vijaya. 2016. "Rejects Reduction in a Retail Bank Using Lean Six Sigma." *Production Planning and Control* 27 (14): 1131–42.

Syntetos, a.a. A., M. Keyes, and M.Z. Z. Babai. 2009. "Demand Categorisation in a European Spare Parts Logistics Network." *International Journal of Operations & Production Management* 29 (3): 292–316.

Syntetos, Aris A., Zied Babai, John E. Boylan, Stephan Kolassa, and Konstantinos Nikolopoulos.

2016. "Supply Chain Forecasting: Theory, Practice, Their Gap and the Future." *European Journal of Operational Research* 252 (1): 1–26.

Tenhiälä, A, and M Ketokivi. 2012. "Order Management in the Customization-Responsiveness Squeeze." *Decision Sciences* 43 (1): 173–206.

Ting, S.L., Y.K. Tse, G.T.S. Ho, S.H. Chung, and G. Pang. 2014. "Mining Logistics Data to Assure the Quality in a Sustainable Food Supply Chain: A Case in the Red Wine Industry." *International Journal of Production Economics* 152 (June): 200–209.

Tsiptsis, Konstantinos, and Antonios Chorianopoulos. 2009. *Data Mining Techniques in CRM: Inside Customer Segmentation*. West Sussex: Wiley.

Visani, Franco, and Monica Bartolini. 2018. "Business Performance Analytics : Exploring the Potential for Performance Management Systems Business Performance Analytics : Exploring the Potential for Performance Management." *Production Planning & Control* 29(1), 51-67.

Visintin, Filippo, Paola Cappanera, Carlo Banditori, and Pamela Danese. 2017. "Development and Implementation of an Operating Room Scheduling Tool: An Action Research Study." *Production Planning & Control* 28 (9). 758–75.

Vitasek, By Kate L, Karl B Manrodt, and Mark Kelly. 2003. "Solving the Supply-Demand Mismatch." *Supply Chain Management Review*, 58–64.

Waller, Matthew a., and Stanley E. Fawcett. 2013. "Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management." *Journal of Business Logistics* 34 (2): 77–84.

Wedel, Michel, and Wagner A. Kamakura. 2000. *Market Segmentation: Conceptual and Methodological Foundations*. 2nd ed. MA: Kluwer Academic Publishers.

Wind, Y. 1978. "Issues and Advances in Segmentation Research." *Journal of Marketing Research (JMR)* 15 (3): 317–37.

Wind, Y, and RN Cardozo. 1974. "Industrial Market Segmentation." *Industrial Marketing Management*, no. 3: 153–66.

Wu, S. David, Berrin Aytac, Rosemary T. Berger, and Chris a. Armbruster. 2006. "Managing Short Life-Cycle Technology Products for Agere Systems." *Interfaces* 36 (3): 234–47.

Xu, Zhenning, Gary L. Frankwick, and Edward Ramirez. 2016. "Effects of Big Data Analytics and Traditional Marketing Analytics on New Product Success: A Knowledge Fusion Perspective." *Journal of Business Research* 69 (5). Elsevier B.V.: 1562–66.

Yankelovich, Daniel, and David Meer. 2006. "Rediscovering Market Segmentation." *Harvard Business Review* 84 (2): 122–31.

Zezzatti, Ortiz, and Carlos Alberto Ochoa. 2012. *Logistics Management and Optimization through Hybrid Artificial Intelligence Systems*. IGI Global.

Zhao, Chuan, Mats Johnsson, and Mingke He. 2017. “Data Mining with Clustering Algorithms to Reduce Packaging Costs: A Case Study.” *Packaging Technology and Science* 30 (5): 173–93.

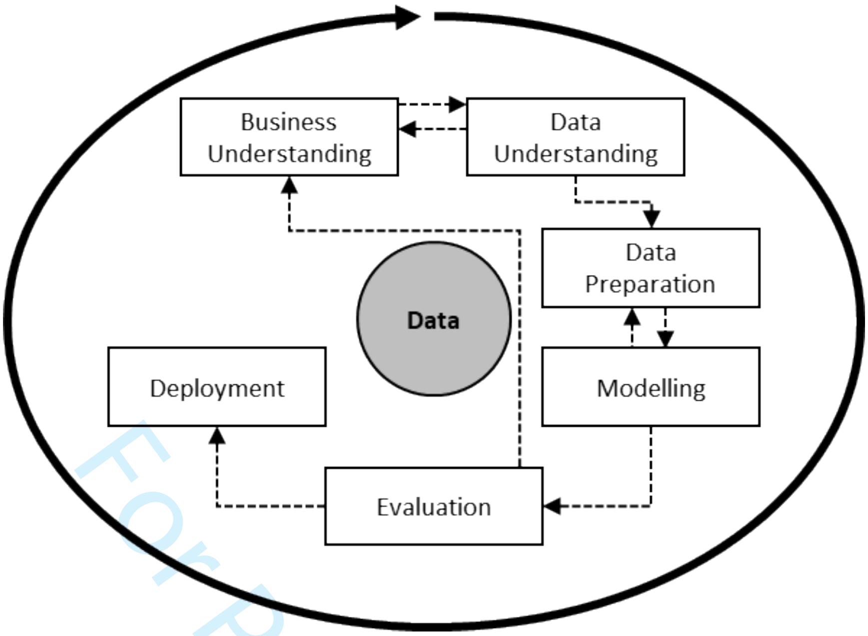

Figure 1 Development process of the framework based in the principles of action research CRISP-DM Cycle

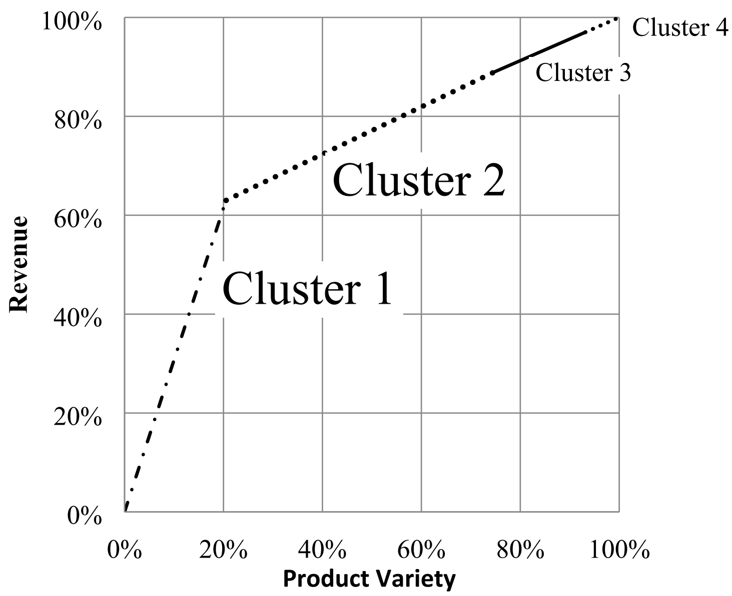

Figure 2: Revenue versus Variety - cumulative analysis of clusters

Table 1 Criteria for Segmentation in SCM literature

Criteria \ Contributions	Oliver and Webber (1982)	Fuller et al., (1993)	Gattorna and Walters (1996)	Fisher (1997)	Naylor *et al.* (1999)	Mason-Jones *et al.* (2000)	Childerhouse and Towill ((2000)	Lamming (2000)	Li & O'Brien (2001)	Christopher & Towill (2002)	Lee (2002)	Childerhouse *et al.* (2002)	Vitasek *et al.* (2003)	Aitken *et al.* (2003)	Bruce *et al.* (2004)	Payne and Peters (2004)	Holweg (2005)	Lovell *et al.* (2005)	Christopher *et al.* (2006)	Christopher *et al.* (2009)	Godsell *et al.* (2011)	Godsell *et al.* (2013)
Demand Variability					✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓
Lead time/Time window	✓		✓	✓		✓	✓	✓		✓	✓	✓		✓			✓		✓	✓		
Product life cycle				✓		✓	✓	✓		✓	✓	✓		✓	✓		✓	✓		✓		
Volume						✓		✓		✓	✓	✓	✓	✓		✓	✓			✓	✓	✓
Product Variety	✓			✓	✓		✓	✓		✓	✓	✓		✓			✓			✓		
Profit margin				✓		✓	✓		✓	✓	✓											
Demand Predictability				✓			✓		✓		✓								✓			✓
Flexibility	✓		✓					✓														
Point of product configuration								✓	✓								✓					
Responsiveness	✓					✓		✓									✓					
Customer expectations						✓		✓	✓								✓					
Reliability of supply	✓		✓					✓			✓											
Product complexity						✓		✓														
De-coupling point						✓											✓					
Frequency of delivery/orders			✓													✓						
Demand Pareto analysis													✓				✓					
Change over	✓										✓											
Range	✓							✓														
Stock out				✓			✓				✓											
Inventory costs				✓			✓				✓											
Minimum run size	✓																					
Quality problems											✓											
Obsolescence											✓											
Number of supply sources											✓											
Number of customers																✓						
Order line value																✓						
Substitutability of SKU																✓						
Price/Revenue		✓																				
Product Value Density																		✓				

Table 2: FMCGCo's characteristics

Characteristic	FMCGCo
Unit of analysis	Company (5 production units)
Time frame under analysis	2 years of operation
Industry	Food processing industry
Business relationship	B2B
Current manufacturing strategy	MTO
Experience	20+ years
Key product	Fruit-based additives and flavourings
Relationship with customers	Strategic supplier
Relationship with suppliers	Large customer (holds leverage)
Product type	Bespoke
Number of SKU	~ 1000
Number of suppliers	~ 250
Number of customers	~ 100
Average capacity utilisation	~75%
Capacity utilisation during high season	~95%
Seasonality	Peak between spring and summer
Demand volume trend	Positive growth
Geographic location	Europe & North of Africa

Table 3: Criteria proposed for collection

Criteria		
Unavailable	**Available**	
Stock outs	Product life cycle	Change over*
Inventory costs	Minimum run size*	Substitutability of SKU
Quality problems	Point of product configuration	Range
Obsolescence	Responsiveness	Frequency of delivery**
Number of supply sources	Customer expectations	Demand Variability**
Order line value	Reliability of supply	Time window for delivery**
Flexibility	Product complexity*	Demand Volume**
Profit margin	Price/Revenue*	Order corrections ratio**
	Demand Predictability*	Frequency of orders
Not applicable to the case	Demand Pareto analysis	Product Variety*
Number of customers		
De-coupling point		
Product value density		

* Criteria considered for PCA analysis **Remaining criteria after PCA iterations

Table 4: Criteria correlation matrix (a=0.01, Sig. 1-tailed)

	Time Window for Delivery	**Order Correction**	**Demand Volume**	**Demand Variability**
Order Corrections	-0.030 (0.184)			
Demand Volume	-0.086 (0.004)	-0.022 (0.255)		
Demand Variability	0.239 (0.000)	0.028 (0.200)	-0.150 (0.000)	
Frequency of deliveries	-0.107 (0.001)	-0.185 (0.000)	0.527 (0.000)	-0.159 (0.000)

Table 5: Total Variance Explained (before rotation)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	Variance %	Cumul. %	Total	Variance %	Cumul. %
1	1.700	33.994	33.994	1.700	33.994	33.994
2	1.135	22.699	56.693	1.135	22.699	56.693
3	0.965	19.300	75.993	0.965	19.300	**75.993**
4	0.753	15.063	91.057			
5	0.447	8.943	100.000			

Table 6: Rotated component matrix (loadings less than 30% omitted)

Criteria	Component			Communalities	
	1	2	3	Initial	Extraction
Demand Volume	0.888			1.000	0.656
Frequency of deliveries	0.846			1.000	0.972
Time Window for Delivery		0.806		1.000	0.802
Demand Variability		0.761		1.000	0.602
Order Corrections			0.984	1.000	0.767

Table 7: Product clusters profiled against the selected criteria

	Criteria	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster profile	Variety	21%	54%	18%	7%
	Revenue	63%	26%	8%	3%
	Demand Volume	74%	10%	8%	8%
	Demand Variability	Stable	n.s.	n.s.	Unstable
	Order Corrections	Average	V. low	High	Low
	Delivery Time Window	V. short	Short	Short	V. long
	Delivery Frequency	Frequent	Sporadic	V. Rare	Average

Table 8: Illustrative matching of SCM planning practices with product clusters

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Operational Priority	Reduce operational cost (Fisher and Raman 1996)	✓			
	Reduce Variety & Stream line processes (Ramdas 2003)		✓		
	Negotiate end of life / substitution (Aitken, Childerhouse, and Towill 2003)			✓	✓
Planning Approach	MTF - Forecast statistically (Syntetos et al. 2016)	✓			
	Collaborative planning/VMI (Lapide 2008)			✓	
	MTO (Tenhiälä and Ketokivi 2012)		✓		✓

Revision Notes

Title: ***Developing a Framework to Support Strategic Supply Chain Segmentation Decisions: a case study***

ID - TPPC-2018-0268.R1 for the Production Planning & Control

Dear Reviewers,

We are very grateful for your insightful comments and suggestions for our revised manuscript "*Developing a Framework to Support Strategic Supply Chain Segmentation Decisions: a case study"* submitted to *Production Planning & Control*.

In our revision, we tried to address all comments and recommendations made. We hope that the revised version of the paper will satisfy reviewers' expectations. In what follows, we provide detailed description of changes we made in response to reviewers' comments. Comments are highlighted in blue italics, our response is typed as normal text and the extracts of the rewritten parts from the manuscript are referenced by page and highlighted in a grey box.

Best Regards

The authors

Reviewer: 1

Recommendation: Revise for new review – lacking industrial relevance or application

Response:

We thank Reviewer 1 for the recommendation and feedback. We have addressed the lack of industrial relevance or application in the revisions detailed below.

Review Comment:

Dear authors

Thank you for submitting this very interesting paper that I was pleased to review. The topic is timely, important and represents potentially a good contribution to the field of SCM. Your paper would be greatly strengthened if you focused more attention on the action research methodology.

I would expect to see the research extend beyond that in the paper in the spirit of action research to engage the practitioners more fully in implementing the segmentation, verifying with them the value of that implementation in terms of their own individual roles and the organisation. Action research is about co-produced research that has high impact but unfortunately the impact story is not told.

I would encourage you to tell this impact story and include quotes from the engaged managers that evidence this impact. This could then be a rich, in-depth, qualitative, action research based single case study.

In terms of writing style, the paper is well written - it is clear, concise and well-structured. However, the writing style deteriorates somewhat towards the end of the paper where the use of many quotes from non-academic sources comes across almost as proselytising. The earlier parts of the paper are crisper and more professionally written.

I wish you well with this paper as I think it has very good potential to make valuable contribution to SCM.

Response:

We are very grateful to Reviewer 1 for the time and effort they have taken to read the paper and thank them for providing valuable comments. We are also very pleased that the reviewer found the paper important and a good contribution to the field of SCM. We have enriched the paper with further detail about the implementation and the impact story. We have added quotes from managers as evidence. We have also removed some of the previous quotes from non-academic sources. Further detail on these changes will be addressed below.

We have made multiple changes in the section on the insights from the case study and now pp.21-23

Insights from the case study

The potential benefits of 'borrowing' from the field of marketing, to mine a blend of structured and unstructured data, against a strategic SC objective have been demonstrated through the FMCGCo case. FMCGCo has the strategic objective to move beyond a single MTO strategy towards a segmented SC strategy and by following the data-mining methodology, the five most relevant criteria (demand volume, demand variability, order corrections, delivery time window and delivery frequency) were identified. These were then used to identify the relevant segments, that require different SC responses. While the primary insight from this study is the empirically tested approach for data mining to support strategic SC decision making, five further specific takeaways for industry emerged.

Finding the Known Unknowns

One of the strengths of this approach to data mining is that it helps to identify the 'known unknowns'. Popularised by Donald Rumsfeld in a Department of Defence briefing (February 12th 2002), 'Known UnKnowns' are things that we know we do not know. Four of the five criteria that were identified are part of Christopher and Towill's (2002) DVW^3 approach. The fifth dimension, order correction, has no previous mention in the literature. It was identified in Step 1, through conversation with the management of FMCGCo. It was part of their business that was known to them, as they had to deal with the impact of changing orders all the time. The significance of this factor was 'unknown' to them and only identified through the data mining.

Turns insight into SC practices

A critical outcome of any form of clustering or segmentation is that it drives differentiated practices within the business. Identifying different segments enables the development of specific marketing programmes for the different segments. Planning is the glue that holds an SC together, and it is not surprising that the primary differentiator between the identified segments in terms of operationalisation is the planning process combined with the operational priority. Building on the first insight, it is 'known' that different SKUs may require different processes, but what can be difficult to determine (i.e. 'unknown') is which SKUs require which process. This method is very effective in identifying the clusters of SKUs that require different planning approaches.

The need to combine data with managerial insight

Managerial insight is critical throughout the process, and particularly in Steps 1 and 3. Although the manager may not have recognised its importance, without the conversation order correction would never have been identified as a potential dimension in Step 1. It is therefore important to ensure that in the first stage of the process, a detailed understanding of the business, its challenges and opportunities are understood to ensure that the initial list of dimensions is as broad as possible. At this stage, it may be useful to involve someone from outside the business, who can look at the business with a fresh pair of eyes so that critical aspects of the SC operation (e.g. customers continually changing their orders) is not overlooked. The operationalisation of the clusters in Step 3 would not be possible without the input of the SC team, who have the SC knowledge to understand the operational priorities and how they can be operationalised in practice.

Understand the relative importance of different segments

George Orwell (1945, p.40) in his novel Animal Farm famously said "All animals are equal, but some animals are more equal than others". The same can be said for SC clusters. As illustrated in Figure 3, Clusters 1 and 2 accounted for over 80% of the revenue, and almost 80% of the variety of the FMCGCo portfolio. These segments were the 'cash cows' for the business. To enable these important segments to be serviced as efficiently as possible, the SKUs in these clusters need to be recognised and given appropriate management attention.

Increasing importance of the SC analyst

While many of the major consultancies are starting to develop their SC capability more formally, and specialist SC analytic companies are still in an embryonic field. As this study has identified, the potential benefits to a company are huge. The question is not about whether to deploy SC analytics in a business, but how best to deploy the approaches. Depending on the size, maturity, and the appetite to put data driven approaches at the heart of the business, companies may choose to develop this as an in-house capability or through engaging the services of a third party.

Review Comment:

Does the motivation for the work originate from industry's current problems? Is the need for the work demonstrated?:

Yes to both questions. The motivation for the work is based on a contemporary and highly topical issue, that of the application of data mining, applied here to supply chain data embedded in a SAP implementation. The authors rightly identify the now common use of data *mining in marketing, notably customer loyalty POS information systems. SCM as a field needs more research using these techniques.*

Response:

Thank you for the positive feedback

Review Comment:

Does the paper describe and evaluate previous literature on the subject?:

Yes. The literature review is thorough. Table 1 provides an excellent summary of the literature relating to each criterion for segmentation of supply chains.

Response:

We are grateful for the very positive comment. We have edited the table 1 with summary of the literature to include the missing Price/Revenue by Fuller et al. (1993) to rectify a comment further in this review.

Review Comment:

Do the authors show awareness of work that has been published recently in PPC?:

Yes, there are a number of recent PPC references within the manuscript. However, in response to a reviewer comment a set of papers published in PPC using single case studies are provided by the authors. These should be incorporated into the methodology section to strengthen the argument for use of a single case.

Response:

We have added the mention of single case studies and action research to the justification of methodology as well as the call for case studies by PPC. Page 11 now reads as follows:

> Whilst a single application does not allow generalisation, it does contribute to the existing body of knowledge on SC segmentation both for academics and practitioners with potential of being used in other cases (e.g., Hofmann et al. 2013; Hersh 2014). Literature offers a wide range of papers based on Single studies and action research (Cagliano et al. 2005; Ferreira, Arantes, and Kharlamov 2015; Farooq and O'Brien 2015; Silva et al. 2016; Sunder 2016; Perona et al. 2016; Visintin et al. 2017; Visani and Bartolini 2019;) and the use of case studies is welcome and encouraged (Childe 2011, 2017).

Review Comment:

Does the paper describe the use of an appropriate research method?:

Unlike a previous reviewer (reviewer 2) I am not personally concerned about a single in-depth case study as part of an action research methodology. However the paper requires strengthening in its description of the methodology and, possibly, in the methodology actually used. The following aspects of the research give rise to methodological concerns:

The authors state the following were involved in initial brainstorming interviews - global SC manager and managers dealing closely with the SAP, production control, production planning and quality management. A table should be provided showing the level of management, duration of the 'interviews', how many researchers were involved, how many researchers made 'contact notes', how were these notes coded and integrated?

Response:

We have added further detail on the problem and setting, quotations by managers and further contextual information. We have opted to include the most interesting and available detail in the narrative rather than a separate table. The paper on pp.12-14 now reads as:

> The project begins with unstructured interviews with managers and brainstorming. The aim is to understand what should be included in the exploratory analysis, understanding expectations and what the limitations, perceived problems and challenges are. Outputs are gathered through contact notes. The team included two researchers and representatives from the company. The company side included the global supply chain manager, production manager and internal control manager all dealing closely with the SAP and involved in production control, production planning and quality management. A total of 15 meetings were held focusing on data collection, development and validation. The average meeting took around two hours; all team members had equal weight in the decision process and the decisions were reached by consensus. When consensus on a decision was impossible to reach the decision was made based on the majority. All of the company sites had an integrated ERP system which enabled easy access to past data for analysis.
>
> The company's overall operational goal and general directive is to be as responsive as possible. This attitude resonates in the global SC manager's own words as: "*…what the client wants, we deliver it. If our client wants star shaped Earliglow strawberries collected at 6 o'clock in the morning on the southern side of a mountain in a specific place of the world, we will engineer a process and find what is needed to deliver him that star shaped Earliglow strawberries…"*
>
> The company works on a pure MTO basis (ETO for new products) which results in a high product variety (about one thousand SKUs). The product quality is very strict and being a perishable good time window for delivery are often narrow. Being a in the middle of the global supply chain, demand unpredictability is a great concern. Minor demand spikes downstream cause classical problems like bullwhip effect (Forrester,

1961), resulting in huge inventory build-ups upstream at the supplier's levels as well as "boom and bust" effect along SC (Sterman, 2000). As managers suggested, this demand variability is the consequence of order and production batching, client's poor inventory management, lack of visibility and information sharing, as well as seasonality and final product marketing strategies (summer promotions). Most of the firms operational efforts focus on coping with the volatile and unstable demand.

One of the problems and long-term concerns expressed by the managers is of production capacity. When discussing the manufacturing utilisation, demand peaks hit 96% utilisation during early summer and considering the current business growth the company was expected to start running out of capacity soon. In order to keep up the productivity without increasing manufacturing capacity, management stressed the importance of levelling the demand.

When discussing potential reasons for demand instability, managers complained about some clients repeatedly correcting their orders causing operational havoc. The team agreed that this could be measured using the ratio of corrected orders of a given SKU divided by the number of orders of the same SKU. Considering the nature of these order corrections, the main corrections were the delay of the delivery date (about 32%), followed by the anticipation of the delivery date (about 19%) and finally corrections to the ordered volume (13%). The team agreed that these corrections were completely driven by clients behaviour and that it causes significant disruptions to planning and manufacturing scheduling, often costing in lost raw materials and poor capacity utilisation.

The project team determined that the overall business goal is to identify the different SKU segments and their characteristics so that FMCGCo can deploy appropriate SC strategies. The SC management objective is to move beyond a one-size fits all strategy towards a segmented strategy. The data-mining objective is therefore to cluster SKUs based on the operational data at the SKU level and to provide management with insights about the characteristics of each cluster informing the SC strategy.

Review Comment:

To facilitate the process a list of criteria previously used for differentiation in the SCM literature is presented (Table 3) and each criterion discussed." The differences between table 1 and table 3 are not fully explained; what happened to product value density; where did revenue come from?

Response:

We have rectified table 1 and table 3 accordingly. The new table 1 read as follows on p.5.

Criteria \ Contributions	Oliver & Webber (1982)	Fuller et al., (1993)	Gattorna & Walters (1996)	Fisher (1997)	Naylor *et al.* (1999)	Mason-Jones *et al.* (2000)	Childerhouse & Towill (2000)	Lamming (2000)	Li & O'Brien (2001)	Christopher & Towill (2002)	Lee (2002)	Childerhouse *et al.* (2002)	Vitasek *et al.* (2003)	Aitken *et al.* (2003)	Bruce *et al.* (2004)	Payne and Peters (2004)	Holweg (2005)	Lovell *et al.* (2005)	Christopher *et al.* (2006)	Christopher *et al.* (2009)	Godsell *et al.* (2011)	Godsell *et al.* (2013)
Demand Variability					✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓
Lead time/Time window	✓		✓	✓		✓	✓	✓		✓	✓	✓		✓			✓		✓	✓		
Product life cycle				✓		✓	✓	✓		✓	✓	✓		✓	✓		✓	✓		✓		
Volume						✓		✓		✓	✓	✓	✓	✓		✓	✓			✓	✓	✓
Product Variety	✓			✓	✓		✓	✓		✓	✓	✓		✓			✓			✓		
Profit margin				✓		✓	✓		✓	✓	✓											
Demand Predictability				✓			✓		✓		✓								✓			✓
Flexibility	✓		✓					✓														
Point of product configuration								✓	✓								✓					
Responsiveness	✓					✓		✓									✓					
Customer expectations						✓		✓	✓								✓					
Reliability of supply	✓		✓					✓			✓											
Product complexity						✓		✓														
De-coupling point						✓											✓					
Frequency of delivery/orders			✓													✓						
Demand Pareto analysis													✓				✓					
Change over	✓										✓											
Range	✓							✓														
Stock out				✓			✓				✓											
Inventory costs				✓			✓				✓											
Minimum run size	✓																					
Quality problems											✓											
Obsolescence											✓											
Number of supply sources											✓											
Number of customers																✓						
Order line value																✓						
Substitutability of SKU																✓						
Price/Revenue		✓																				
Product Value Density																		✓				

And Table 3 on p.14 reads as follows:

Table 3: Criteria proposed for collection

Criteria		
Unavailable	**Available**	
Stock outs	Product life cycle	Change over*
Inventory costs	Minimum run size*	Substitutability of SKU
Quality problems	Point of product configuration	Range
Obsolescence	Responsiveness	Frequency of delivery**
Number of supply sources	Customer expectations	Demand Variability**
Order line value	Reliability of supply	Time window for delivery**
Flexibility	Product complexity*	Demand Volume**
Profit margin	Price/Revenue*	Order corrections ratio**
	Demand Predictability*	Frequency of orders
Not applicable to the case	Demand Pareto analysis	Product Variety*
Number of customers		
De-coupling point		
Product value density		

* Criteria considered for PCA analysis **Remaining criteria after PCA iterations

Review Comment:

Does the paper provide useful new knowledge about the application of the work in practice or to provide directions for future research?:

Not currently. My main concern with the paper is the limited implementation of the action research methodology. Engaged scholarship of any type is highly valuable because it should generate findings that evidence impact. The research looks half finished in the way it is presented. How were the findings verified and discussed with the practitioner partners to the co-produced research? Were the findings of the segmentation exercise implemented? What were the benefits of the implementation? I would have expected a substantial discussion involving quotes from each of the managers involved about how this research made a difference to them in their roles and to the company as a whole. Otherwise, what is the point of using an action research methodology?

Response:

This comment was partly addressed in the previous changes with further detail on the problem and setting, quotations by managers and further contextual information on pp.12-14 and more specifically regarding the implementation and implications we have added a new section on pp.23-24 and a new paragraph on and expanded the paragraph in the Conclusions section on p.24. The new parts now read as follows respectively:

Implementation and implications for practice

After the deployment stage the firm spent the following months implementing the insights of this study. The follow up evaluation of benefits of this study was not scoped, but the team made a qualitative assessment of the implementation. It was common agreement that we cannot fully attribute all the improvements in firms operational performance observed over the following months to the changes made based on this study, however, the team believes that it had a crucial role.

Part of the produce is now MTF rather than MTO. A new programme started aiming at establishing VMI and collaborative planning with its main client. Some of the SKUs were discontinued and substitutes common to several clients were developed and agreed with clients motivated by a lower price due to the economy of scale. Following these changes, the firm observed a noticeable reduction in operational cost and increase in productivity though better capacity utilisation due to demand levelling and reduction of variability. Some of the “bad clients” were identified and contacted with suggestions to address the order corrections problem.

The global SC manager says that: “… *we are still offering the best in class customisation and most of what we produce is MTO because that is still the core of what we do. But we no longer do this for every single client or product, part is standardized which allows us to have lower costs and we share this benefit with our clients. It’s a clear win-win for all*!” adding that “…*we realise that by trying to respond to any customer demand we were underperforming. Now we are looking for a healthy balance between being responsive and being efficient*.”

Regardless of all the positive dynamic observed in a relatively short time-frame, the team realises that maximizing the potential of SC segmentation will take a long time, effort and commitment. The close involvement of both researchers and practitioners though the action research cycle was also praised due to its agility and efficiency and benefitted all parties involved.

Expanded conclusion paragraph on p.24:

The proposed framework aims at integrating data mining methods into supply chain segmentation. By applying the CRISP-DM Cycle and Action Research we identify the five most relevant criteria (demand volume, demand variability, order corrections, delivery time window and delivery frequency). We use these criteria to identify the four segments that require different SC responses. The action research used for this study contributes to developing new knowledge as well as achieving practical results. The main advantages of action research is the close involvement of both researchers and practitioners in a common effort. The cyclical process of planning, acting and evaluating is the core of action research methodology, enabling a much faster research process compared to other, more traditional and detached methods. We identify the following five key takeaways: finding the Known Unknowns, turns insight into SC practices, the need to combine data with managerial insight, understand the relative importance of different clusters and increasing importance of the SC analyst.

The framework is innovative due to the application of data mining methods in the SCM context. The framework is based on The Cross Industry Standard Process for Data Mining. This makes our framework scalable and compatible with much larger portfolios. The main characteristics of the framework is its open-ended nature, expansibility and adaptability. The use of exploratory data mining methods not only allows validation but also discovering new insights from data. The proposed framework can be used as a regular diagnostic check tool to constantly re-assess segments based on the most recent demand history partly solving the problem of demand dynamics.

Review Comment:

Do you think the paper has sufficient promise to make revisions worthwhile?:

Definitely, yes. The topic is very interesting, the methodology is also of interest.

Response:

We are grateful for the positive feedback.

Reviewer: 2

Recommendation: Accept as-is or minor revisions? no further review

Response:

We are very grateful for the positive feedback. Some improvements have been implements as suggested. Further detail is provided below.

Review Comment:

The authors have addressed the concerns of the reviewers in an appropriate manner. The use of literature snowballing is noteworthy. I would like to see it explained so other researchers can benefit from the approach.

Response:

We included a brief explanation and reference to the snowballing method and the paragraph now reads as follows on p.4:

> At the current state, SC segmentation is normative and close-ended suggesting very specific criteria to be used. Different authors suggest different criteria depending on the context and purpose, and there are many possibilities suggested in the literature as listed in Table 1 which has been created using literature snowballing (Sayers 2007) and is in accordance with criteria mentioned throughout Protopappa-Sieke and Thonemann (2017). Literature Snowballing is a technique borrowed from systematic reviews as per Sayers (2007) and it is essentially working with the papers cited in the most recent literature identified and retrieving them, then lifting the citations from those until no more criteria have been identified.

Review Comment:

Does the motivation for the work originate from industry's current problems? Is the need for the work demonstrated?:

Deciding what criteria to involve in segmentation continues to be an issue for industry. This paper provides a framework supported by analytics which can aid managers in deriving and updating segments.

Response:

We are grateful for the positive feedback.

Review Comment:

Does the paper describe and evaluate previous literature on the subject?:

The revised paper covers the literature however, the authors need to ensure that the listing of papers by date is consistent.

Response:

We have rectified the listing of papers so it is consistent by date.

On p.6 the references now read as follows:

(e.g. Vitasek et al. 2003; Payne and Peters 2004; Holweg 2005; Godsell et al. 2011)

On p.6 the references now read as follows:

(Dreyer et al. 2016; Datta 2017). Tailored practices (Lapide 2005a, 2005b; Godsell et al. 2011)

On p.7 the references now read as follows:

(Wind and Cardozo 1974; Green 1977; Plank 1985; Henry and William 1986; Yankelovich and Meer 2006)

On p.7 the references now read as follows:

(Wind 1978; Plank 1985; Wedel and Kamakura 2000; Kazbare, van Trijp, and Eskildsen 2010)

On p.10 the references now read as follows:

(Shearer et al. 2000; Mariscal, Marbán, and Fernández 2010; Zezzatti and Ochoa 2012; Shafique and Qaiser 2014)

Review Comment:

Do the authors show awareness of work that has been published recently in PPC?: yes

Response:

We are grateful for the positive feedback.

Review Comment:

Does the paper describe the use of an appropriate research method?:

This is a strong part of the paper. Missing is a description of "literature snowballing" as a technique to derive relevant criteria. The paper would be enhanced by the authors providing a short explanation of the technique as it is not commonly used.

Response:

We are grateful for the positive feedback and as previously mentioned we have added an explanation to the literature snowballing and a reference to a paper that provides further detail on the method. The paragraph on p.4 now read as follows:

At the current state, SC segmentation is normative and close-ended suggesting very specific criteria to be used. Different authors suggest different criteria depending on the context and purpose, and there are many possibilities suggested in the literature as listed in Table 1 which has been created using literature snowballing (Sayers 2007) and is in accordance with criteria mentioned throughout Protopappa-Sieke and Thonemann (2017). Literature Snowballing is a technique borrowed from systematic reviews as per Sayers (2007) and it is essentially working with the papers cited in the most recent literature identified and retrieving them, then lifting the citations from those until no more criteria have been identified.

Review Comment:

Does the paper provide useful new knowledge about the application of the work in practice or to provide directions for future research?:

The paper contributes to the on-going segmentation debate

Response:

We are grateful for the positive comment.

Review Comment:

Do you think the paper has sufficient promise to make revisions worthwhile?:

The revisions required are minor

Response:

We appreciate the positive feedback and we have implemented further revisions to address all the suggestions.